NCER National Centre for Econometric Research

NCER Working Paper Series

Estimating Stochastic Volatility Models Using a Discrete Non-linear Filter

A. Clements, S. Hurn and S. White

Working Paper #3

August 2006

Abstract

Many approaches have been proposed for estimating stochastic volatility (SV) models, a number of which are filtering methods. While non-linear filtering methods are superior to linear approaches, non-linear filtering methods have not gained a wide acceptance in the econometrics literature due to their computational cost. This paper proposes a discretised non-linear filtering (DNF) algorithm for the estimation of latent variable models. It is shown that the DNF approach leads to significant computational gains relative to other procedures in the context of SV estimation without any associated loss in accuracy. It is also shown how a number of extensions to standard SV models can be accommodated within the DNF algorithm.

Estimating Stochastic Volatility Models Using a Discrete Non-linear Filter

Adam Clements, Stan Hurn, Scott White.
School of Economics and Finance, Queensland University of Technology

July 25, 2006

Abstract

Many approaches have been proposed for estimating stochastic volatility (SV) models, a number of which are filtering methods. While non-linear filtering methods are superior to linear approaches, non-linear filtering methods have not gained a wide acceptance in the econometrics literature due to their computational cost. This paper proposes a discretised non-linear filtering (DNF) algorithm for the estimation of latent variable models. It is shown that the DNF approach leads to significant computational gains relative to other procedures in the context of SV estimation without any associated loss in accuracy. It is also shown how a number of extensions to standard SV models can be accommodated within the DNF algorithm.

Keywords
non-linear filtering, stochastic volatility, state-space models, asymmetries, latent factors, two factor volatility models.

JEL Classification C13, C22, C53

Corresponding author
Adam Clements
School of Economics and Finance
Queensland University of Technology
GPO Box 2434 Brisbane, 4001
Qld, Australia
Ph: +61 7 3864 2525

email `a.clements@qut.edu.au`
The authors wisk to thank seminar participants at the Queensland University of Technology, the University of Technology Sydney and The 2004 Australasian Econometric Society meetings. We would also like to thank Adrian Pagan and Gael Martin for comments on the paper. Any errors or omissions are of course the responsibility of the authors.

1 INTRODUCTION

The stochastic volatility (SV) class of models has proved particularly useful in capturing the time-varying volatility of financial asset returns. This popularity has spawned a large literature on methods for estimating the parameters of SV models. These include: Quasi Maximum Likelihood (Harvey, Ruiz and Shephard, 1994), Generalized Method of Moments (Melino and Turnbull, 1990), Efficient Method of Moments (Gallant and Tauchen, 1996), Simulated Maximum Likelihood (Danielsson and Richard, 1993; Danielsson 1994), Monte-Carlo Maximum Likelihood (Sandman and Koopman, 1998) and a number of Bayesian procedures that use MCMC (Jacquier *et al.* 1994; Kim, Shephard and Chib, 1998; Chib, Nardari and Shephard 2002). A full (as opposed to quasi) maximum likelihood procedure that does not rely on simulation requires application of the nonlinear filtering framework introduced by Kitagawa (1987).

Despite the generality of Kitigawa's algorithm, it has not been widely adopted in the empirical literature. In their comment on Kitagawa (1987), Martin and Raferty (1987) argued that the computational cost of the proposed numerical integration procedure is so great that the method was unlikely to be of practical use, a sentiment echoed by Ghysels, Harvey and Renault (1996). Indeed, only Fridman and Harris (1998) and Watanabe (1999) have used Kitigawa's algorithm in the SV context. The major contribution of this paper is the development of a discrete non-linear filtering (DNF) algorithm for the evaluation of Kitigawa's set of non-linear filtering equations, and hence a computationally-feasible maximum likelihood method for the estimation of the parameters of SV models.

The DNF is based on a fixed discretisation of the state-space of the latent factor(s), thus allowing continuously-valued latent-variables to be dealt with as if they were discrete-valued Markov processes. Monte Carlo simulations show that this approach allows significant reduction in the computational cost of maximum likelihood estimation, without any concomitant reduction in the efficiency of the parameter estimates. The flexibility of the DNF algorithm is demonstrated by using it to estimate three non-standard SV specifications,

namely, the heavy-tailed, asymmetric, and two-factor SV models.

The remainder of the paper is structured as follows. Section 2 outlines the general non-linear filtering framework found in the work of Kitigawa (1987). Section 3 sets out the proposed DNF method. In Section 4 the basic SV framework is outlined together with details of how the DNF estimation procedure is applied to this class of model. This section also contains results of a Monte-carlo experiment to highlight the efficacy of the DNF algorithm. Section 5 outlines how the DNF estimation algorithm can accommodate extensions to the standard SV model, specifically heavy tails, leverage and multiple volatility factors. In Section 6 the standard and extended SV specifications are applied to a series of S&P500 returns. Section 7 provides concluding remarks.

2 THE NON-LINEAR FILTERING FRAMEWORK

Consider a system described by the state-space model

$$y_t \sim r\ (.|\, x_t, Y_{t-1,}\, \theta), \quad x_t \sim q\ (.|x_{t-1}, Y_{t-1,}\, \theta) \tag{1}$$

where y_t is an observed data series conditional on the value of the (unobserved) state variable x_t, Y_{t-1} represents all observable information up to and including time $t-1$ and θ is an unknown the parameter vector to be estimated. In this representation, $r\ (.|\, x_t, Y_{t-1,}\, \theta)$ is the conditional likelihood of y_t given the state variable x_t, and $q\ (.|x_{t-1}, Y_{t-1,}\, \theta)$ is the transition probability distribution of x_t given x_{t-1}. In the event of that $r\ (.|\, x_t, Y_{t-1,}\, \theta)$ and $q\ (.|x_{t-1}, Y_{t-1,}\, \theta)$ are linear functions and with $y_t \sim N(0, \sigma_u^2)$ and $x_t \sim N(0, \sigma_w^2)$, standard linear Kalman filtering techniques may be used to generate maximum likelihood estimates of the unknown parameters, θ (see Harvey, **?**). In the more general case where linearity or normality does not apply, the maximum likelihood estimates of θ are given by

$$\begin{aligned} \widehat{\theta}_{ML} &= \arg\max \left[f\ \left(\{y_t\}_{t=1}^{T} \,|\, \theta \right) \right] \\ &= \arg\max \left[\int .. \int f\ \left(\{y_t\}_{t=1}^{T} \,|\, \{x_t\}_{t=1}^{T} \right)\ f\ \left(\{x_t\}_{t=1}^{T} \,|\, \theta \right)\ dx_1 .. dx_T \right] \end{aligned} \tag{2}$$

which is a T dimensional integration problem.

A general approach to the problem of evaluating the integral in equation (2) is provided by the recursive prediction-update algorithm suggested by Kitagawa (1987). As in equation (1), let $r\left(y_t|\, x_t, \theta\right)$ be the conditional distribution of y_t on x_t and $q\left(x_t|\, x_{t-1}, \theta\right)$ be the conditional distribution of x_t on x_{t-1}. The one-step ahead prediction of the distribution of x_t conditional on Y_{t-1}, $f\left(x_t\,|Y_{t-1}, \theta\right)$, is given by

$$f\left(x_t\,|Y_{t-1}, \theta\right) = \int_{-\infty}^{\infty} q(x_t\,|x_{t-1}, Y_{t-1},\ \theta)\, f\left(x_{t-1}\,|Y_{t-1}, \theta\right) dx_{t-1}. \tag{3}$$

Once a new observation, y_t, is available, the probability distribution of the state variable at time t, conditional on information at time t, $f\left(x_t\,|Y_t, \theta\right)$, may now be obtained as

$$f\left(x_t|\, Y_t, \theta\right) = \frac{r(y_t|x_t, Y_{t-1}, \theta)\, f\left(x_t\,|Y_{t-1}, \theta\right)}{f\left(y_t|\, Y_{t-1}, \theta\right)}. \tag{4}$$

The denominator of equation (4) is the likelihood of the observation y_t conditional on Y_{t-1} and θ and may be computed as

$$f\left(y_t|\, Y_{t-1}, \theta\right) = \int_{-\infty}^{\infty} r(y_t|x_t, Y_{t-1}, \theta)\, f\left(x_t|\, Y_{t-1}, \theta\right) dx_t. \tag{5}$$

There are two important by-products obtained by recursion through equations (3) and (4) for all observations T. In the first instance the log-likelihood function used to generate ML estimates of θ is obtained directly from equation (5) and is given by

$$\ln L = \sum_{t=1}^{T} \ln[f\left(y_t|\, Y_{t-1}, \theta\right)]. \tag{6}$$

In addition to parameter estimation, recursion of the filter allows the smoothed distribution of x, conditional on all information up to and including T to be determined. Note that the distribution of x_t conditional on Y_T and θ is constructed as

$$f\left(x_t|\, Y_T, \theta\right) = f\left(x_t|\, Y_t, \theta\right) \int_{-\infty}^{\infty} \frac{f\left(x_{t+1}|\, Y_T, \theta\right) q(x_{t+1}\,|x_t, Y_T,\ \theta)}{f\left(x_{t+1}\,|Y_t, \theta\right)}\, dx_{t+1}, \tag{7}$$

with expected value

$$E\left[x_t|\, Y_T, \theta\right] = \int_{-\infty}^{\infty} x_t \cdot f\left(x_t|\, Y_T, \theta\right) dx_t. \tag{8}$$

From the perspective of parameter estimation it is clear that the intractable high-dimensional integral in equation (2) has been replaced with the relatively straightforward summation in equation (6). The problem, of course, is to provide a numerical technique to evaluate the integrals in the prediction and update equations, (3) and (4). Kitagawa (1987) suggests that the relevant integrals be evaluated using trapezoidal integration which leads to the *pdf* of the state variable being approximated by a piecewise-linear spline. This requires the specification of the number of linear segments in the spline, the location of the spline knots and consequently the value of the functions, that is the heights of the probability densities $f\left(x_t \left| Y_{t-1}, \theta\right.\right)$ and $f\left(x_t \right| Y_t, \theta)$, at the knots[1]. As pointed out by Martin and Raferty (1987), the piecewise spline procedure is computationally very demanding to implement.The next section, therefore, is devoted to the description of an alternative approach that delivers significant computational gains without any deterioration in numerical accuracy.

3 THE DISCRETE NON-LINEAR FILTER

The discrete non-linear filter (DNF) is based on a discretisation of the state-space of a continuous latent variable. This allows the likelihood function in equation (2) to be evaluated in a manner similar to that used for Markov models of discrete valued time series (see MacDonald and Zucchini, 1997). This avoids the use of numerical integration schemes. In the DNF algorithm the probability density function (pdf) of the latent variable, x, is approximated by computing the probability of observing x within a set of discrete intervals. This discretisation is based on defining N adjacent intervals in x space, bounded by $w^1 \ldots w^{N+1}$, and centered on the points $x^1 ... x^N$ where

$$x^i = \frac{w^i + w^{i+1}}{2}. \tag{9}$$

The probability of observing $x \in (w^i, w^{i+1}]$, that is x is within the interval centered on x^i, is given by

$$p(x \in (w^i, w^{i+1}]) = \int_{w^i}^{w^{i+1}} f\left(x\right) dx \approx p(x^i) \tag{10}$$

[1]Kitagawa (1987) proposed a very simple scheme for knot placement with knots equally spaced over the finite interval taken to be the domain of the state variable.

where $f(x)$ is the pdf of the of the unobserved state variable x. The values $\left\{p(x^i)\right\}_{i=1}^{N}$ constitute a discrete approximation to the continuous distribution $f(x)$. Both the prediction and update distributions from equations (3) and (4) will be constructed in this way.

Since the DNF is based on discrete approximations, the transition distribution of x, $q(x_t\,|x_{t-1}, Y_{t-1},\ \theta)$, may be thought of in terms of transition probabilities. An $N \times N$ transitional probability matrix, $\widehat{q}$, is defined whose elements $(\widehat{q}^{\,i,j}\ \forall i,j = 1,...,N)$ represent the probability of x migrating from the interval centred on x^j to the interval centred on x^i. The elements of $\widehat{q}$ are constructed as

$$\widehat{q}^{\,i,j} = p(x_t \in (w^i, w^{i+1})\,|\,x_{t-1} \in (w^{\,j}, w^{\,j+1}), Y_{t-1},\ \theta) \approx \delta\ q\left(x^i|\,x^j, Y_{t-1},\ \theta\right) \tag{11}$$

where $q(.)$ is the transition probability distribution of x and δ is the interval width. Similarly, the likelihood of observing y_t conditional on x_t is also approximated for each discrete interval. The $N \times 1$ conditional likelihood vector, $\widehat{r_t}$, has elements $\widehat{r}\,_t^{\,i}$ given by

$$\widehat{r}\,_t^{\,i} = r(y_t\,|\,x \in (w^i, w^{i+1}), Y_{t-1},\ \theta) \approx r\left(y_t|\,x^i, Y_{t-1},\ \theta\right) \qquad i = 1\ldots N. \tag{12}$$

where, as before, $r(.)$ is the conditional likelihood function of y_t upon x_t, Y_{t-1} and θ.

After defining the transition matrix and the conditional likelihoods, the DNF proceeds with the following steps.

Prediction

In the general nonlinear filtering case, the distribution of the latent variable is predicted using equation (3). In the DNF, $f(x_t\,|Y_{t-1}, \theta)$ is replaced by the discrete approaximation, $\left\{P_t^i\right\}_{i=1}^{N}$, representing the predicted probability that x is an element of each discrete interval at time t. This prediction is given by

$$\begin{aligned} P_t^i &\approx \sum_{j=1}^{N} \delta\, q(x_t^i\,|x_{t-1}^j, Y_{t-1},\ \theta)\, p\,(x_{t-1}^j\,|Y_{t-1}, \theta) \\ &= \sum_{j=1}^{N} \widehat{q}^{\,i,j} \cdot U_{t-1}^j \end{aligned} \tag{13}$$

where U_{t-1}^{j} it the time $t-1$ updated probability that x lies within the j^{th} interval. Since $\{P_t^i\}_{i=1}^N$ must constitute a proper *pdf*, it's elements must sum to one. Since there may be slight approximation error in constructing $\{P_t^i\}_{i=1}^N$, it must be standardised to sum to one.

Likelihood

Given $\{P_t^i\}_{i=1}^N$, the continuous integral required to evaluate the likelihood function in equation (5) is now evaluted as follows:

$$\begin{aligned} f\left(y_t|\, Y_{t-1},\theta\right) &\approx \sum_{i=1}^{N} r(y_t|x_t^i, Y_{t-1},\ \theta) p\left(x_t^i|\, Y_{t-1},\theta\right) \\ &= \sum_{i=1}^{N} \widehat{r}_{\ t}^{\ i} \cdot P_t^i. \end{aligned} \tag{14}$$

Update

After y_t has been observed, equation (4) is used to update the distribution of x. The DNF uses U_t^i to represent the updated probability that x lies within the i^{th} interval:

$$\begin{aligned} U_t^i &= \frac{r(y_t|x_t^i, Y_{t-1},\theta)\, p\left(x_t^i|Y_{t-1},\theta\right)}{f\left(y_t|\, Y_{t-1},\theta\right)} \\ &= \frac{\widehat{r}_{\ t}^{\ i} \cdot P_t^i}{f\left(y_t|\, Y_{t-1},\theta\right)} \end{aligned} \tag{15}$$

The update for the i^{th} interval is simply the weighted conditional likelihoods normalised by the overall likelihood. This set of updated probabilities, $\{U_t^i\}_{i=1}^N$ represents a discrete approximation to the continuous update distribution in equation (4).

Fixed Interval Smoothing

Fixed interval smoothing provides a method for generating estimates of the expected value of the state variable conditional upon all available information. Given a value for θ, the smoothed distribution of x_t conditional on information up to and including time T can be generated. The smoothed probability that x lies with the i^{th} interval at time t is denoted $S_t^i = p(x_t^i\,|\,Y_T,\theta)$ and is found by

$$S_t^i = U_t^i \cdot \sum_{j=1}^{N} \frac{P_{t+1}^j \cdot \widehat{q}^{\,i,j}}{S_{t+1}^j}. \tag{16}$$

This smoothing procedure works backward through the data and begins by setting $S_T^i = U_T^i$. The smoothed value of the latent variable is then simply found as:

$$E\left[x_t \middle| Y_T, \theta\right] = \sum_{i=1}^{N} x^i \cdot S_t^i. \tag{17}$$

Initialisation

For the DNF to be initialized, the prediction of the state probabilities at time $t = 1$ need to be selected. The state probabilities are initialized by discretising the unconditional distribution of the state variable such that

$$P_1^i = \int_{w^i}^{w^{i+1}} f(x \,|\, \theta)\, dx. \tag{18}$$

where $f(x \,|\, \theta)$ is the unconditional distribution of x given the elements in the parameter vector θ.

While the DNF procedure is designed for use in a wide range of latent variable problems, this paper will now examine its application to the stochastic volatility class of models.

4 ESTIMATING THE STANDARD STOCHASTIC VOLATILITY MODEL

The discrete time stochastic volatility (SV) model introduced by Taylor (1982, 1986) specifies the returns of a financial asset as:

$$y_t = \sigma_t u_t \qquad u_t \sim N\left(0, 1\right) \tag{19}$$

where σ_t is the time t conditional standard deviation of y_t. The returns $\{y_t\}_{t=1}^T$ are an observed variable, but the model treats σ_t as an unobserved (latent) stochastic variable. The simplest SV model specifies $\ln(\sigma_t^2)$ as an AR(1) process,

$$x_t = \alpha + \beta\, x_{t-1} + w_t \qquad w_t \sim N(0, \sigma_w^2) \tag{20}$$

where $x_t = \ln(\sigma_t^2)$ and $E[u_t,w_t] = 0$.

To implement the DNF the first step is to define a set of intervals bounded by $w^1 \ldots w^{N+1}$. For SV estimation purposes, points are chosen to be uniformly

distributed in x ($\ln(\sigma^2)$) space to span

$$\frac{\alpha}{(1-\beta)} \pm C\frac{\sigma_w}{\sqrt{(1-\beta^2)}}$$

from which N discrete intervals centred on $x^1 ... x^N$ are defined which span C standard deviations each side of the unconditional mean. For all subsequent empirical work, $N = 25$ or 50 and $C = 6$. Simulation studies subsequently discussed in this section examine the issue of interval placement.

For the DNF to be applied, the conditional likelihood function for y_t and the transition distribution for x_t must be defined. From equation (19) the conditional likelihood function is

$$r(y_t \,|\, x_t, Y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi \exp(x_t)}} \exp\left[-\frac{y_t^2}{2\exp(x_t)}\right]. \tag{21}$$

The elements of the conditional likelihood vector, $\widehat{r}_t$ are then found as:

$$\widehat{r}_t^{\,i} = \frac{1}{\sqrt{2\pi \exp(x^i)}} \exp\left[-\frac{y_t^2}{2\exp(x^i)}\right]. \tag{22}$$

From equation (20), the transition distribution of x_t is found as

$$q(x_t \,|\, x_{t-1}, Y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left[-\frac{(x_t - \alpha - \beta\, x_{t-1})^2}{2\sigma_w^2}\right]. \tag{23}$$

The definition of the transition distribution leads to the elements of the transition probability matrix being

$$\widehat{q}^{\,i,j} = \frac{\delta}{\sqrt{2\pi\sigma_w^2}} \exp\left[-\frac{(x^i - \alpha - \beta\, x^j)^2}{2\sigma_w^2}\right]. \tag{24}$$

To obtain the initial profile of $\left\{P_1^i\right\}_{i=1}^N$, equation (18) is used setting

$$f\,(x|\,\theta) \sim N\left(\frac{\alpha}{(1-\beta)}, \frac{\sigma_w^2}{(1-\beta^2)}\right).$$

Given $\widehat{q}^{\,i,j}$, $\widehat{r}_t^{\,i}$, and P_1^i, equation (13) through (15) are used to evaluate the log-likelihood of $\{y_t\}_{t=1}^T$.

Based on $\widehat{\theta}_{ML}$, the expected value of unobserved volatility may be extracted from the smoothed distribution, $p_t(x \in (w^i, w^{i+1}] \,|\, y_T, \widehat{\theta}_{ML})$ constructed using equation (16),

$$E(\sigma_t^2|\, Y_T, \widehat{\theta}_{ML}) = \sum_{i=1}^{N} \exp(x_i)\, p(x_t \in (w^i, w^{i+1}] \,|\, Y_T, \widehat{\theta}_{ML}). \tag{25}$$

It is clear from the description of the DNF, that under the discretisation scheme, the number of intervals chosen and interval placement are important issues. The results of two simulation studies reported here give some reasonable guidance in terms of selecting the number of intervals and how to distribute them in state space. These results support the choice of $C = 6$ and intervals of equal width. These studies utilise the Monte-Carlo framework proposed by Jacquier *et al.* (1994) which considered the the three parameter sets,

$$\begin{aligned}\theta_1 &= (\alpha,\, \beta,\, \sigma_w) = (-0.736,\, 0.90,\, 0.363),\\ \theta_2 &= (\alpha,\, \beta,\, \sigma_w) = (-0.368,\, 0.95,\, 0.260),\\ \theta_3 &= (\alpha,\, \beta,\, \sigma_w) = (-0.147,\, 0.98,\, 0.166).\end{aligned}$$

The first issue considered is related to the choice of N and C, the importance of which will be examined in the context of accuracy of likelihood evaluation. As discussed in Section 2, the true likelihood of latent variable models such as SV models is an intractable high dimensional integral. To investigate the impact of the choice of N and C in terms of accuracy of likelihood evaluation a benchmark is required. For this purpose, the benchmark is the likelihood obtained from an SV model using the DNF given arbitrarily large values for N and C. Values of $N = 500$ and $C = 10$ are chosen with this likelihood being referred to as $\widehat{L}_{10,500}$ below. Given this benchmark, the impact on the accuracy of likelihood evaluation will be highlighted by considering how well the DNF procedure approximates $\widehat{L}_{10,500}$ by using a range of smaller values for N and C. The DNF procedure will be applied using all combinations of $C = \{3, 4, 5, 6, 8, 10\}$ and $N = \{25, 50, 75, 100, 500\}$. For each combination, RMSE is computed between the estimated likelihood and $\widehat{L}_{10,500}$ given 1000 replications of a sample size of $T = 2000$. Results are reported using θ_1, θ_2 and θ_3.

The results of this simulation, reported in Table 1 reveal a number of interesting patterns. Clearly as either N or C decrease, the accuracy with which the SV likelihood is evaluated relative to $\widehat{L}_{10,500}$ decreases. It is evident however that the choice of C is more important in that it has a larger impact on accuracy relative to reductions in N. The rate at which accuracy of likelihood evaluation deteriorates rises rapidly when $C < 6$, irrespective of the choice of

			DNF					
	N	C	3	4	5	6	8	10
	25		1.6377	1.4244	0.1143	0.0050	0.0052	0.0110
	50		1.4501	1.2554	0.0939	0.0026	0.0013	0.0024
θ_1	75		1.3985	1.2005	0.0874	0.0022	0.0006	0.0011
	100		1.3596	1.1734	0.0841	0.0020	0.0003	0.0006
	500		1.2888	1.1092	0.0765	0.0017	0.0000	–
	25		2.2349	2.2695	0.3338	0.0316	0.0172	0.5352
	50		2.0130	2.0415	0.2904	0.0251	0.0017	0.0030
θ_2	75		1.9393	1.9668	0.2761	0.0229	0.0007	0.0013
	100		1.9026	1.9297	0.2690	0.0218	0.0004	0.0007
	500		1.8149	1.8416	0.2522	0.0192	0.0001	–
	25		2.2067	2.1582	0.1605	0.5414	2.2956	22.029
	50		1.9996	1.9440	0.1351	0.0018	0.0021	0.0436
θ_3	75		1.9241	1.8672	0.1258	0.0012	0.0009	0.0016
	100		1.8856	1.8280	0.1210	0.0010	0.0005	0.0009
	500		1.7910	1.7326	0.1093	0.0008	0.0000	–

Table 1: Accuracy of likelihood estimation for the DNF and NFML procedures. For each parameter set, tabled here are the RMSE for the DNF procedure with various combinations of N and C, along with the RMSE for the NFML procedure with various N. To provide RMSE figures, an estimate of the true likelihood is taken to be the DNF procedure with 500 intervals spanning 10 standard deviations on each side of the unconditional mean of the latent variable.

N, indicating that $C = 6$ ensures the intervals adequately span the state-space. Thus to balance computation time with accuracy, $C = 6$ is chosen for all applications of the DNF. Reducing N to relatively small values such as 50 does not appear to have a dramatic effect on accuracy, thus for all empirical application considered, $N = 50$ is chosen as it represents a reasonable point in the trade-off between accuracy and computational cost.

Now the issue of interval distribution is addressed, in doing so two robust conclusions arise. Table 2 compares the performance of the DNF procedure to the NF (nonlinear filter) algorithm of Kitigawa (1987) and the NFML approach of Watanabe (1999). Like the NFML, the NF procedure utilises a trapezoidal integration scheme, however, in the NF procedure a fixed grid of points is

	N	DNF	NF	$NFML$
	25	0.0050	0.0038	2.034
θ_1	50	0.0026	0.0019	0.8173
	75	0.0022	0.0016	0.492
	100	0.0020	0.0014	0.3292
	25	0.0316	0.0398	2.741
θ_2	50	0.0251	0.0286	0.898
	75	0.0229	0.0257	0.516
	100	0.0218	0.0243	0.338
	25	0.5414	0.4369	4.453
θ_3	50	0.0018	0.0020	1.057
	75	0.0012	0.0015	0.584
	100	0.0010	0.0013	0.340

Table 2: RMSE in likelihood evaluation of competing nonlinear filtering approaches. The DNF is compared to the trapezoidal approach (NF) and the NFML. For both the DNF and NF procedures C is chosen to be 6, the NFML procedure is implemented as in Watanabe (1999).

chosen in the same manner as in the DNF. Since the grid is fixed, the transition function between points needs only to be evaluated once. Motivated by the previous findings, only the results for $C = 6$ are reported for both the DNF and NF procedures. The NFML algorithm is implemented exactly as in Watanabe (1999). For each estimation procedure the RMSE is calculated by comparing the estimated likelihood values with $\widehat{L}_{10,500}$ from the DNF algorithm.

First, by comparing the results for the DNF and NF procedures it is clearly seen that the discrete approximation used in the DNF is of comparable accuracy to the more computationally burdensome NF trapezoidal integration approach. For both procedures the RMSE is very similar for each choice of N. For the third parameter set, the NF and DNF procedures lead to similar degrees of inaccuracy when $N = 25$. This indicates that the loss in accuracy for this choice of N is not a problem specific to the DNF procedure. Second, when the results of the NFML algorithm are considered it is clear that this procedure is very inaccurate. This inaccuracy arises from the discretisation scheme employed

in the NFML whereby random, normally distributed, nodes are selected based on the output from the Kalman Filter. By choosing normally distributed nodes it is conjectured that accuracy in the tails of the distribution is sacrificed for resolution about the expected value of the latent variable. This, coupled with inaccuracy of the Kalman filter in the SV setting, leads to the possibility that the majority of nodes are placed in the wrong region of state-space.

Overall, these results indicate that the discrete (DNF) approach proposed here provides a relatively accurate method for estimating the likelihood of latent variable models. Along with this accuracy, the DNF is computationally cheap. Average computation time for one parameter evaluation using the DNF scheme being dramatically lower in comparison to more complex algorithms. For example, when $T = 2000$ the average time for one parameter evaluation using the DNF is 7.21 seconds, compared with 454 seconds for the NFML algorithm. To provide a reference time, the QML procedure takes an average of 2.43 seconds[2].

Now the performance of the DNF will be considered in the *SV* parameter estimation context, where its performance will be directly compared to alternative SV estimators. Specifically, the nonlinear filtering procedures of Fridman and Harris (denoted FH) (1998) and Watanabe (denoted NFML) (1999) and the MCMC procedure of Jacquier *et al.* (1994) will be considered. Once again, the Monte-Carlo framework of Jacquier *et al.* (1994) is utilised, with series of lengths $T = 500$, and $T = 2000$ bieng simulated from equations (19) and (20), and the parameter estimates obtained from these series are stored. The process is repeated 1000 times. Table 3 reports the mean and root mean squared error (RMSE) for estimates using the DNF.

A number of conclusions emerge from the results reported in Table 3. The major result is that the DNF procedure produces comparable results to the Bayesian estimator of Jacquier et al. (1994). Furthermore, the DNF procedure exhibits comparable results to the NFML and FH algorithms which are based on numerical integration of the nonlinear filtering equations.

[2]Times recorded for MatLab® code using a common minimisation routine, on a a Pentium IV 2.8GhZ desktop computer.

	$T = 500$				$T = 2000$			
	DNF		FH	$MCMC$	DNF		$NFML$	$MCMC$
N	(25)	(50)			(25)	(50)	(50)	
$\alpha = -0.736$	-0.881	-0.881	-0.87	-0.87	-0.765	-0.765	-0.776	-0.762
	(0.385)	(0.385)	(0.43)	(0.34)	(0.163)	(0.159)	(0.168)	(0.15)
$\beta = 0.900$	0.881	0.881	0.88	0.88	0.896	0.896	0.895	0.896
	(0.052)	(0.052)	(0.05)	(0.046)	(0.022)	(0.021)	(0.023)	(0.02)
$\sigma_w = 0.363$	0.379	0.379	0.37	0.35	0.365	0.364	0.368	0.359
	(0.081)	(0.081)	(0.08)	(0.067)	(0.040)	(0.041)	(0.041)	(0.034)
$\alpha = -0.368$	-0.500	-0.496	-0.51	-0.56	-0.397	-0.395	-0.406	–
	(0.298)	(0.299)	(0.306)	(0.34)	(0.098)	(0.100)	(0.106)	
$\beta = 0.950$	0.932	0.933	0.93	0.92	0.946	0.946	0.945	–
	(0.040)	(0.041)	(0.04)	(0.046)	(0.013)	(0.013)	(0.014)	
$\sigma_w = 0.260$	0.275	0.273	0.28	0.28	0.264	0.263	0.264	–
	(0.065)	(0.066)	(0.07)	(0.065)	(0.030)	(0.031)	(0.032)	
$\alpha = -0.147$	-0.273	-0.254	-0.09	-0.22	-0.201	-0.169	-0.178	–
	(0.207)	(0.206)	(0.09)	(0.14)	(0.068)	(0.058)	(0.067)	
$\beta = 0.980$	0.963	0.965	0.987	0.97	0.973	0.977	0.976	–
	(0.028)	(0.028)	(0.015)	(0.02)	(0.009)	(0.008)	(0.009)	
$\sigma_w = 0.166$	0.195	0.179	0.18	0.23	0.189	0.169	0.169	–
	(0.057)	(0.053)	(0.04)	(0.08)	(0.030)	(0.022)	(0.024)	

Table 3: Simulation results for the DNF, Fridman and Harris (FH), MCMC, and NFML procedures. For each parameter set, the mean parameters and RMSE (in brackets) are reported. N denotes the number of intervals and nodes used in the DNF and NFML respectively.

Given the first two parameter combinations, reducing the number of intervals from $N = 50$ to $N = 25$ does not significantly impact on the performance of the DNF. This is in contrast to the NFML procedure where decreasing the number of nodes from $N = 50$ to $N = 25$ results in reduced accuracy[3]. It is conjectured that this difference in performance relates to the placement of nodes/intervals. Equally spaced intervals trade off resolution near the unconditional mean of x for increased resolution in the tails.

To conclude, the DNF does not require numerical integration to approximate the likelihood of a latent variable process. The Monte Carlo results indicated that the discrete approximation does not impact adversely on accuracy but delivers significant reduction in computational cost. Having established the DNF procedure as a viable SV estimation procedure, its use in estimating the parameters of a number of extended SV specifications is now discussed.

[3] See Table I in Watanabe (1999).

5 EXTENSIONS

This section considers three extensions to the standard SV model and how the basic DNF framework may be modified to accommodate each of them. Section 5.1 reveals how the DNF may be modified to incorporate non-normal error distributions into a standard SV model, thus permitting a heavy-tailed SV model to be estimated. Section 5.2 shows how the DNF can accommodate correlation between return and volatility innovations, an important feature when dealing with equity returns to capture the leverage effect. Section 5.3 considers how the DNF methodology may be applied to dealing with a two factor variance process.

5.1 SV and Heavy Tails

Apart from the issue of modeling time variation in volatility, much research has focused on the shape of the conditional distribution of returns, specifically whether it is non-normal. In the context of SV models, the possibility of non-normal errors implies a more general specification of equation 19

$$y_t = \sigma_t \; u_t \qquad u_t \sim i.i.d.(0,1). \tag{26}$$

In comparison to the standard SV model where $u_t \sim N(0,1)$, a choice relating to the distribution governing u_t must be made. Chib *et al.* (2002) and Jacquier *et al.* (2004) examine the case where u_t is drawn from a $student-t$ distribution using MCMC estimation procedures. Liesenfeld and Jung (2000) and Watanabe and Asai (2001) consider the case where u_t may be drawn from a generalised error distribution (GED) using SML and MCMC respectively.

Either $student{-}t$ or GED error distributions can incorporated into the DNF framework. The only change to the algorithm of Section 4 is to redefine the conditional likelihood distribution, $r(y_t \,|\, x_t, Y_{t-1}, \theta)$. In the current context, the standardised $student-t$ distribution is utilised (model denoted as the $SV-t$ model). Given this choice, $r(y_t \,|\, x_t, Y_{t-1}, \theta)$ is defined as

$$r(y_t \,|\, x_t, Y_{t-1}, \theta) = [\pi(v-2)\exp(x_t)]^{-\frac{1}{2}} \frac{\Gamma((v-1)/2)}{\Gamma(v/2)} \left[1 + \frac{{y_t}^2}{\exp(x_t)(v-2)}\right]^{-\frac{v+1}{2}} \tag{27}$$

where v is the degrees of freedom (that now becomes an extra parameter to be estimated). For the purposes of implementing the DNF, the likelihood vector of equation (22) is redefined as

$$\widehat{r}_t^i = [\pi(v-2)\exp(x^i)]^{-\frac{1}{2}} \frac{\Gamma((v-1)/2)}{\Gamma(v/2)} \left[1 + \frac{{y_t}^2}{\exp(x^i)(v-2)}\right]^{-\frac{v+1}{2}}. \qquad (28)$$

Estimation of the $SV-t$ model simply follows the steps outlined in Section 3.

To assess the ability of the DNF procedure to estimate the parameters of the $SV-t$ model, one parameter set considered in Section 4 is extended to include three different degrees of freedom. The parameter set for the variance equation is $\{\alpha, \beta, \sigma_w\} = \{-0.147, 0.98, 0.166\}$ which is chosen as it is seen to reflect the variance dynamics of daily return (Jacquier *et al*, 1994). The degrees of freedom are chosen to be $v = \{6, 8, 12\}$ resulting in levels of kurtosis of 6, 4.5, and 3.75 respectively.

The simulation study is conducted by simulating a series of length 2000 from the $SV-t$ model with the parameters estimated using the DNF. This process is repeated 1000 times for each parameter set. The mean and RMSE for each parameter set is then found. Following Liesenfeld and Jung (2000) the degrees of freedom is referred to in terms of $1/v$. To assess the impact of the number of intervals selected, this simulation study is carried out for both $N = 25$ and $N = 50$ intervals. The results of this study can be found in Table 4.

Examining the results of the DNF as applied to the $SV-t$ model it is apparent that the DNF accurately estimates the three variance parameters, $\{\alpha, \beta, \sigma_w\}$. The mean and RMSE in estimation of these three parameters is virtually identical to those seen in the estimation of the standard SV model (see Table 3). The only point of minor concern are the slight upward biases in the estimation of α (for both $N = 25$ and 50) and in σ_w (for $N = 25$). This however is consistent with the results for the standard SV model as discussed in Section 4.

The pleasing result is the accuracy with which the DNF procedure estimates the degrees of freedom parameter. From Table 4 it is evident that the DNF procedure as applied to the $SV-t$ model exhibits virtually no bias in estimating

		$SV-t$					
		α	β	σ_w	$1/v$	$1/v$	$1/v$
		-0.147	0.98	0.166	0.166	0.125	0.083
$N=25$	t_6	-0.188 (0.065)	0.973 (0.010)	0.188 (0.032)	0.164 (0.026)		
	t_8	-0.190 (0.062)	0.973 (0.009)	0.186 (0.029)		0.123 (0.025)	
	t_{12}	-0.192 (0.063)	0.973 (0.009)	0.186 (0.029)			0.080 (0.026)
$N=50$	t_6	-0.163 (0.059)	0.977 (0.009)	0.171 (0.027)	0.165 (0.024)		
	t_8	-0.166 (0.062)	0.976 (0.009)	0.170 (0.025)		0.123 (0.025)	
	t_{12}	-0.167 (0.058)	0.978 (0.008)	0.170 (0.025)			0.081 (0.026)

Table 4: Simulation results for the SV-t model with 1000 simulated series of lenght T=2000. The parameter set α, β, σ_w is augmented to include three degrees of freedom, 6, 8, 12. Mean parameter estimates are reported with RMSE in brackets. Estimation is conducted for both N=25 and N=50 intervals.

$1/v$ for either $N = 25$ or $N = 50$ intervals. Furthermore, the RMSE is low and consistent across the three degrees of freedom considered here.

To provide comparative results for the estimation of the $SV - t$ model, the simulation study of Chib *et al.* (2002) is repeated. Here 1000 series of length $T = 1500$ and $T = 3000$ are simulated from the $SV - t$ model with the parameters $\{\alpha, \beta, \sigma_w, v\} = \{-0.15, 0.985, 0.12, 8\}$. Following Chib *et al.* (2002) the sampling properties of $\mu = \frac{\alpha}{(1-\beta)}$ are reported in the place of α, and v instead of $1/v$.

Comparing the variance parameter estimates it is clear that the DNF generates slightly more accurate parameter estimates. This increase in accuracy is most evident in the reduction of bias in the σ_w parameter for $T = 1500$. A surprising result is that the standard deviation of the estimated μ parameter increases with the sample size for the MCMC procedure. Both these results must be taken with some caution due to the low number of simulations (50 replications) used by Chib *et al.* (2002). Examining the estimates of the degrees of

		$T = 1500$		$T = 3000$	
	True	DNF	MCMC	DNF	MCMC
μ	-10	-10.00 (0.198)	-10.00 (0.230)	-10.00 (0.145)	-9.97 (0.272)
β	0.985	0.980 (0.010)	0.976 (0.012)	0.983 (0.0059)	0.981 (0.0055)
σ_w	0.12	0.128 (0.027)	0.1446 (0.024)	0.122 (0.017)	0.123 (0.021)
v	8	8.84 (3.11)	9.66 (4.42)	8.33 (1.44)	8.81 (1.32)

Table 5: Simulation results for the $SV - t$ model with 1000 simulated series of length T=1500 and T=3000. Mean parameter estimates are reported with standard deviation in brackets. The parameter set and MCMC results are replicated from Table 3 in Chib et al (2002). Following Chib et al (2002), mean and RMSE figures are not given for α but for $\mu = \alpha(1\text{-}\beta)^{-1}$. DNF estimation is conducted for N=50 intervals. It is noted that Chib et al (2002) used 50 relications not 1000.

freedom parameter, v, the DNF procedure produces more efficient estimates for both $T = 1500$, and less bias for $T = 3000$.0

5.2 SV and Leverage

With respect to equity returns, Black (1976) and Campbell and Hentschel (1992) theoretically justify the presence of negative correlation between returns and volatility innovations. Generally speaking this feature of equity returns has become known as the leverage effect. Harvey and Shephard (1996) developed an asymmetric SV (ASV) model based on the QML which incorporates the leverage effect by allowing for correlation between return and volatility innovations. Both Yu (2005) and Sandman and Koopman (1998) have proposed ASV models using MCMC and MCL methods respectively. The ASV specification considered here is

$$\begin{aligned} y_t &= \sigma_t u_t \qquad u_t \sim N(0,1) \\ \ln \sigma_t^2 &= \alpha + \beta \ln \sigma_{t-1}^2 + w_t \qquad w_t \sim N(0, \sigma_w^2) \\ \rho &= E[u_{t-1}, w_t]. \end{aligned} \tag{29}$$

To estimate such a specification using the DNF, the standard filtering techniques must be modified to accommodate the correlation, ρ. Given this correlation, the transitional density of equation (23) must be augmented as the evolution of volatility is dependant upon past return observations in the following manner.

$$\begin{aligned} q(x_t|\, x_{t-1},\, y_{t-1}, Y_{t-2}) &= \frac{\delta}{\sqrt{2\pi\sigma_v^2(1-\rho^2)}} \exp\left(-\frac{(x_t - z_t)^2}{2\sigma_v^2(1-\rho^2)}\right), \qquad (30) \\ z_t &= \alpha + \beta\, x_{t-1} + \sigma_v \rho y_{t-1} \exp\left(\frac{-x_{t-1}}{2}\right). \end{aligned}$$

Given $\rho \neq 0$, dependence upon y_{t-1} requires a time-varying transition matrix $\widehat{q}_t$. Elements of this matrix, $\widehat{q}_t^{\,i,j}$ represent the probability that x migrates from interval j to interval i between the distinct times $t-1$ and t respectively. The elements of this matrix are denoted by

$$\widehat{q}_t^{\,i,j} = \delta q(x^i|\, x^j,\, y_{t-1}, Y_{t-2})$$

where δ is the interval width. From equation (30), the elements of the transition matrix are found as:

$$\begin{aligned} \widehat{q}_t^{\,i,j} &= \frac{\delta}{\sqrt{2\pi\sigma_v^2(1-\rho^2)}} \exp\left(-\frac{\left(x_t^i - z_t^j\right)^2}{2\sigma_v^2(1-\rho^2)}\right) \qquad (31) \\ z_t^j &= \alpha + \beta\, x^j + \sigma_v \rho y_{t-1} \exp\left(\frac{-x^j}{2}\right). \end{aligned}$$

Upon computing the series of transition probabilities, the remainder of the DNF is once again unchanged.

To examine the ability of the DNF algorithm to capture the traditional leverage effect, in the form of $\rho < 0$, two simulation experiments have been conducted. The first examines the performance of the DNF approach in isolation given various value of ρ. The second experiment examines the performance of the DNF relative to existing procedures.

Table 6 reports the simulation results for the ASV model estimated using the DNF for values of $\rho = 0,\ -0.3,\ -0.5$ and -0.7. It is clear from these results that applying the DNF algorithm to the ASV estimation leads to accurate estimates of ρ regardless of its magnitude. The associated SV parameters

ASV Model	α $-.363$	β 0.95	σ_v 0.26	ρ_1
$\rho_1 = 0$	-0.40 (0.098)	0.946 (0.013)	0.265 (0.030)	0.0003 (0.079)
$\rho_1 = -0.3$	-0.396 (0.092)	0.946 (0.0125)	0.263 (0.029)	-0.306 (.073)
$\rho_1 = -0.5$	-0.386 (0.082)	0.948 (0.011)	0.261 (0.026)	-0.502 (0.064)
$\rho_1 = -0.7$	-0.379 (0.063)	0.949 (0.085)	0.261 (0.023)	-0.704 (0.052)

Table 6: Simulation Results for the DNF applied to the ASV1 model. 500 replications for a simulated series length of 2000 are conducted. Mean parameter estimates are reported with RMSE in brackets

continue to be reliably estimated after the inclusion of ρ. It should be noted that estimates of α appear to be marginally downward biased, a pattern also observed with the standard SV model in Table 3 irrespective of the estimation procedure used.

Results in Table 7 allow for comparisons to be drawn between the performance of the QML, MCMC, and DNF approaches in relation to the estimation of the ASV model. In terms of ρ, it is clear that the DNF approach produces superior estimates, exhibiting the least bias and RMSE. While there is little difference in relative performance in terms of mean values of β, the RMSE of DNF estimates are lower than the competing approaches. In relation to $\log(\sigma_w^2)$, while there is little to discriminate between the approaches in relation to the mean of the estimates, the DNF does lead to marginally lower RMSE.

5.3 Two-Factor SV

While the standard SV model considered in Section 4 is based on the premise that one latent factor determines the evolution of conditional volatility, two factor SV ($SV2$) models have met with empirical success (Alizadeh *et al.*, 2002 and Liesenfeld and Richard, 2003). This approach allows for the dynamics of conditional volatility to be governed by two independent factors. In practice, it seems as though one factor is very persistent and controls the overall level of

	ρ	β	$\log(\sigma_w^2)$
True	-0.90	0.975	-4.605
QML	-0.911 (0.079)	0.974 (0.007)	-4.617 (0.353)
MCMC	-0.8815 (0.0445)	0.9732 (0.0050)	-4.595 (0.2086)
DNF	-0.9013 (0.0359)	0.9753 (0.0042)	-4.606 (0.2108)

Table 7: Simulation Results for the DNF, and MCMC, and QML procedures applied to the ASV model. The number of simulated series for the DNF procedure is 500 with the sample length being 3000. Mean estimates are reported along with RMSE in parenthese.

volatility, while the second is not persistent and relatively noisy. It has been argued that such a factor structure links transitory shocks to volatility and the tail behavior of the return distribution (Chernov *et al.*, 2003).

A possible specification of an $SV2$ model is

$$\begin{aligned} y_t &= \exp\left(\frac{x_t}{2}\right)\varepsilon_t \\ x_t &= \alpha + x_{1,t} + x_{2,t} \\ x_{1,t} &= \beta_1 x_{1,t} + \sigma_{w,1}\epsilon_{1,t} \\ x_{2,t} &= \beta_2 x_{2,t} + \sigma_{w,2}\epsilon_{2,t} \end{aligned} \tag{32}$$

where ε_t, $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are uncorrelated $N(0,1)$ innovations.

From equation (32), it is seen that the $SV2$ model contains two latent variables. To evaluate the likelihood function conditioned upon the two latent factors, a modified filtering procedure must be used. This procedure must capture the evolution of both x_1 and x_2 through time and the allow for the likelihood to be dependant upon x_1 and x_2. To account for this, the standard likelihood function of equation (21) is now defined as

$$r(y_t|x_{1,t}, x_{2,t}, Y_{t-1}) = \frac{1}{\sqrt{2\pi\exp(\alpha + x_{1,t} + x_{2,t})}} \exp\left(\frac{-y_t^2}{2\exp(\alpha + x_{1,t} + x_{2,t})}\right). \tag{33}$$

Two independent transition distributions governing $x_{1,t}$ and $x_{2,t}$ must be must

also be defined

$$q(x_{1,t}|x_{1,t-1}, Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_{v,1}^2}} \exp\left(\frac{-\left(x_{1,t} - \beta x_{1,t-1}\right)^2}{2\sigma_{v,1}^2}\right), \tag{34}$$

$$q(x_{2,t}|x_{2,t-1}, Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_{v,2}^2}} \exp\left(\frac{-\left(x_{2,t} - \beta x_{2,t-1}\right)^2}{2\sigma_{v,2}^2}\right). \tag{35}$$

As the $SV2$ requires the density of two factors to be integrated through time, two sets of intervals must be chosen such that they span the state-space of $x_{1,t}$ and $x_{2,t}$. The first set of intervals are denoted as $\{w_1^i\}_{i=1}^{N+1}$, and are defined such that they span 6 standard deviations either side of $E[x_1]$. From equation (32), $E[x_1] = 0$ and $V[x_1] = \frac{\sigma_{w,1}^2}{1-\beta_1^2}$. The centers of these intervals are denoted as $x_1^i = \frac{w_1^i + w_1^{i+1}}{2}$. In a similar fashion, the second set of intervals are denoted as $\{w_2^m\}_{m=1}^{N+1}$, and are defined such that they span 6 standard deviations on each side of the unconditional mean of $E[x_2]$. From equation (32), $E[x_2] = 0$ and $V[x_2] = \frac{\sigma_{w,2}^2}{1-\beta_2^2}$. The centers of these intervals are denoted as $x_2^m = \frac{w_2^m + w_2^{m+1}}{2}$.

Given this discretisation, the set of conditional likelihoods for observation y_t is an $N \times N$ matrix denoted as $\widehat{r}_t$. The elements of this matrix are $\widehat{r}_t^{\,i,m}$ and represent the likelihood of y_t given x_1 is within interval i and x_2 is within interval m,

$$\widehat{r}_t^{\,i,m} = r\left(y_t|\, x_t^i,\, x_t^m,\, Y_{t-1},\, \theta\right). \tag{36}$$

From equation (33), $\widehat{r}_t^{\,i,m}$ is found by

$$\widehat{r}_t^{\,i,m} = \frac{1}{\sqrt{2\pi\exp(\alpha + x_1^i + x_2^m)}} \exp\left(\frac{-y_t^2}{2\exp(\alpha + x_1^i + x_2^m)}\right). \tag{37}$$

To apply the DNF to the $SV2$ model requires the definition of two transition matrices $\widehat{q_1}$ and $\widehat{q_2}$, governing $x_{1,t}$ and $x_{2,t}$ respectively. The elements of $\widehat{q_1}$ and $\widehat{q_2}$ represent

$$\begin{aligned} \widehat{q_1}^{\,i,j} &= \delta_1 q(x_1^i|\, x_1^j, Y_{t-1}) \\ \widehat{q_2}^{\,m,n} &= \delta_2 q(x_2^m|\, x_2^n, Y_{t-1}) \end{aligned} \tag{38}$$

where δ_1 and δ_2 are the respective interval widths for each discretisation. From equations (34) and (35) , the elements of these transition matrices are given by

$$\begin{aligned} q_1^{i,j} &= p(x_1^i|x_1^j, Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_{w,1}^2}} \exp\left(\frac{-\left(x_1^i - \beta_1 x_1^j\right)^2}{2\sigma_{w,1}^2}\right), \qquad (39) \\ q_2^{m,n} &= p(x_2^m|x_2^n, Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_{w,2}^2}} \exp\left(\frac{-\left(x_2^m - \beta_2 x_2^n\right)^2}{2\sigma_{w,2}^2}\right). \end{aligned}$$

Based on these definitions, the DNF proceeds as follows.

One-step ahead predictions of the distribution of x_1 and x_2 are found by

$$\begin{aligned} P_{1,t}^i &= \sum_{j=1}^{N} q_1^{i,j} \cdot U_{1,t-1}^j, \qquad (40) \\ P_{2,t}^m &= \sum_{n=1}^{N} q_2^{m,n} \cdot U_{2,t-1}^n. \end{aligned}$$

At the initial time step, these are initialised given their unconditional distributions, $f(x_1|\theta) \sim N(0, \sigma_{v,1}^2/(1-\beta_1^2)$ and $f(x_2|\theta) \sim N(0, \sigma_{v,2}^2/(1-\beta_2^2)$. Since $P_{1,t}$ and $P_{2,t}$ must represent legitimate probability distributions, their elements must are standardised such that they sum to one to eliminate any approximation error.

The likelihood of each observation at time t is found as

$$p(y_t|Y_{t-1}) = \sum_{i=1}^{N}\sum_{m=1}^{N} r_t^{i,m} \cdot P_{1,t}^i \cdot P_{2,t}^m. \qquad (41)$$

Each new observation is incorporated into the updated distribution as follows

$$U_{1,t}^i = \frac{\sum_{m=1}^{N} r_t^{i,m} \cdot P_{1,t}^i \cdot P_{2,t}^m}{p(y_t|Y_{t-1})}, \qquad (42)$$

$$U_{2,t}^i = \frac{\sum_{i=1}^{N} r_t^{i,m} \cdot P_{1,t}^i \cdot P_{2,t}^m}{p(y_t|Y_{t-1})}. \qquad (43)$$

A simulation study is undertaken to examine the accuracy with which the DNF procedure estimates the $SV2$ parameters. Series of length $T = 2000$ are simulated from the $SV2$ model in equation (32) (the true parameters are reported in Table 8), with the parameters estimated using the DNF methodology.

	$Model: SV2$				
Parameter	α	β_1	$\sigma_{w,1}$	β_2	$\sigma_{w,2}$
Actual	0	0.98	0.1	0.4	0.8
DNF	−0.010 (0.121)	0.963 (0.046)	0.126 (0.082)	0.367 (0.122)	0.782 (0.096)

Table 8: Simulation results for the SV-t model with 1000 simulated series of length T=1500 and T=3000. Mean parameter estimates are reported with RMSE in brackets. The parameter set and MCMC results are replicated from Table 3 in Chib et al (2002). Following Chib et al (2002), mean and RMSE figures are not given for α but for $\mu = \alpha(1\text{-}\beta)^{-1}$. DNF estimation is conducted for N=50 intervals

This procedure is repeated 1000 times. The results of this simulation study are outlined in Table 8.

Examining the results for the two β parameters it is seen that the DNF procedure produces slight downward biases for both parameters. On examination of the parameter estimates of the simulated samples it is clear that this downward bias stems from only a few parameter estimates that are well below the target values. Examining the RMSE of the parameter estimates reveals that the parameters of the first volatility factor (high persistence) are estimated more accurately than that of the second factor (low persistence). This pattern is most evident in the estimation of the β parameters. Here, the RMSE for the second factor is three times that of the first factor. This is to be expected as the signal produced by the first factor is very strong and thus more readily identifiable.

6 EMPIRICAL APPLICATION

This section applies the DNF algorithm to generating one step ahead predictions of S&P500 volatility. This analysis is based on daily returns from the S&P500 index spanning 2 January 1990 to 16 August 2004 (3689 observations). The full sample of 3689 observations is split into an estimation period containing the first 2689 observations and a hold-out sample of the final 1000 observa-

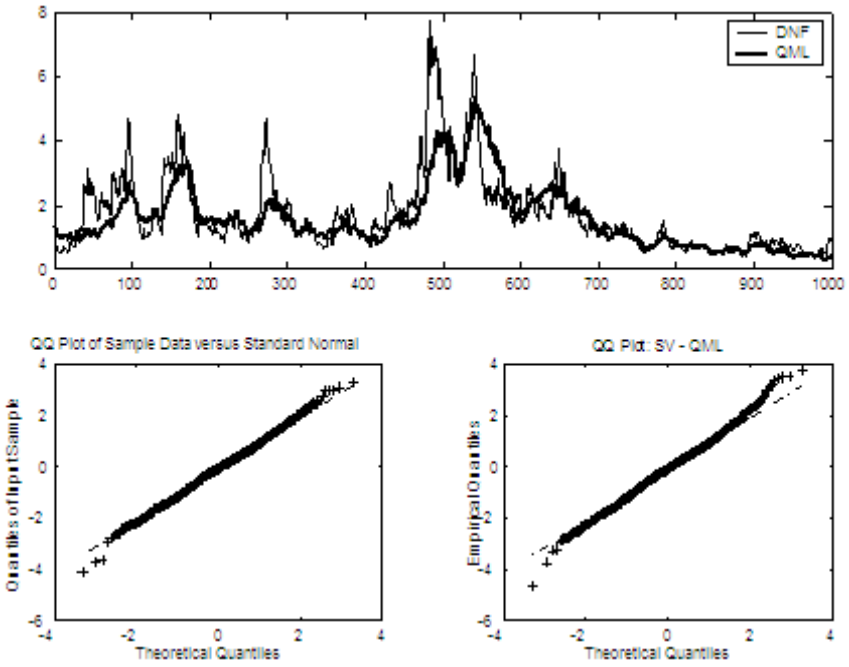

Figure 1: Comparison of QML and DNF one-step ahead predictions (top panel) based on the hold-out sample. QQ-plots of returns standardised by one-step ahead DNF volatility predictions (bottom left panel) and one-step ahead QML volatility predictions given hold-out sample.

tions. A comparison of the standard and extended SV models will be based on returns from the hold out sample using parameters estimated from the estimation period. Comparisons will be drawn by examining QQplots of standardised residuals and the out of sample log likelihood[4].

To highlight the overall benefit of using nonlinear filtering in the context of the standard SV model, the DNF is compared to the QML procedure. Figure 1 plots the DNF and QML one-step ahead predictions for the hold-out sample (top panel) and the respective QQ-plots of returns standardised by the one-step ahead predictions of volatility (bottom panel). The top panel indicates that the DNF adapts more quickly to general changes in the level of S&P 500 volatility. In many instances, the DNF predictions rise (fall) somewhat earlier than the corresponding QML predictions. By comparing the QQ-plots in the bottom

[4]The out of sample log likelihood is calculated as $\sum_{t=T+1}^{T+1000} \log(f(y_t|Y_{t-1}, \widehat{\theta}_{ML}))$. Where $\widehat{\theta}_{ML}$ is the maximum likelihood estimate of θ from returns in the estimation period $\{y_t\}_{t=1}^{T}$.

	Model			
Parameter	*SV*	$SV-t$	*ASV*	*SV2*
α	-0.004 (0.003)	-0.003 (0.003)	-0.016 (0.005)	-0.420 (0.219)
β_1	0.986 (0.004)	0.985 (0.002)	0.977 (0.004)	0.984 (0.002)
$\sigma_{w,1}$	0.131 (0.019)	0.133 (0.020)	0.185 (0.017)	0.133 (0.019)
$1/v$		0.125 (0.021)		
ρ			-0.580 (0.060)	
β_2				-0.139 (0.152)
$\sigma_{w,2}$				0.576 (0.057)
Likelihood				
Insample	-3466.3	-3444.9	-3444.7	-3444.1
Hold Out Sample	-1651.8	-1657.4	-1633.8	-1656.6

Table 9: Insample parameter estimates for the SV, $SV-t$, ASV and $SV2$ models for the SP500 return series with associated standard errors in parentheses. For ease of comparison to the $SV2$ model, β and σ_w for the SV, $SV-t$ and ASV models have been relabled β_1 and $\sigma_{w,1}$ respectively. Likelihood values have been given for both the insample estimation period and the out of sample forecast evaluation period.

panels of Figure 1, it becomes clear that the ability of the DNF to quickly adapt to changes in volatility results in superior forecasts.

Having seen that the DNF provides improvements over the QML procedure, the focus now turns to the incremental effect of extended SV specifications, relative to the standard SV model. Table 9 reports the results for the DNF applied to the standard SV, $SV-t$, ASV and $SV2$ models. In-sample results from the $SV-t$ model indicate that after accounting for time-varying volatility, S&P 500 returns are conditionally non-normal,.$1/v = 0.125$ or $v = 8$. The significance of this feature is reflected in the in-sample likelihood ratio statistic[5] of $LR_{1/v=0} = 42.8$, $(\chi^2_{1,\ 0.05} = 3.841)$. Furthermore, allowing for heavy-tails

[5]Likelihood ratio statistics of the three extended SV specifications are determined by comparing the log likelihoods of the respective models to that of the standard SV model.

does not influence the estimates of the three variance parameters.

The parameter estimates of the ASV model highlight the importance of allowing return and variance innovations to be correlated. The importance of the ρ coefficient in this case is confirmed by the high estimated value $\rho = -0.580$ leading to a significant increase in the likelihood ($\text{LR}_{\rho=0} = 43.2,\ \chi^2_{1,.05} = 3.841$). By incorporating correlation, the estimated value of β is marginally decreased with a slight increase seen in the estimate of σ_w.

Estimation results for the $SV2$ model indicate that adding a second volatility factor provides a significant increase in likelihood $LR_{\beta_2=\sigma_{w,2}=0} = 44.4,\ (\chi^2_{2,.05} = 5.991)$. Examining the parameter estimates, it is seen that two dramatically different factors are driving changes in volatility. Factor 1 is found to exhibit a high degree of persistence ($\beta_1 = 0.984$) and a low level of noise ($\sigma_{w,1} = 0.133$). Conversely, factor 2 has a negative persistence parameter ($\beta_2 = -0.139$) and a very high level of noise ($\sigma_{w,2} = 0.574$). It could be conjectured at this stage that the second volatility factor is simply noise that proxies for a misspecified return distribution.

Out of sample volatility plots of the four SV models are contained in Figure 2. Examining the top and bottom panels reveals a distinct similarity between the variance estimates generated by the $SV - t$ and $SV2$ models. Both the $SV - t$ and $SV2$ models produce volatility estimates that that are slower to respond than the respective SV predictions. Conversely, taking into account the leverage effect, leads to the ASV producing volatility predictions that rise and fall marginally faster than the SV model, as shown in the middle panel of Figure 2.

The relative accuracy of the out of sample volatility estimates can be ascertained by examining the QQ-plots of standardised residuals in Figure 3 and the out of sample likelihood values in Table 9. The first panel of Figure 3 reveals that the weakness of the SV model is in capturing extreme returns. The associated out of sample likelihood (-1651.8) is the benchmark to which the three extended models will be compared.

The similarity of the volatility predictions from the $SV - t$ and $SV2$ models

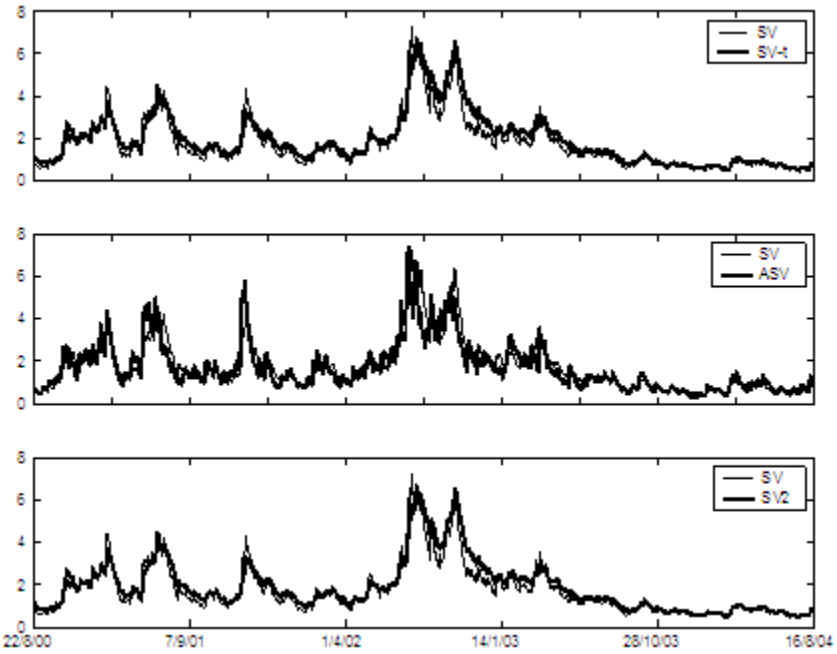

Figure 2: Conditional volatility series from the out of sample period. $SV-t$, ASV and $SV2$ volatilities are shown realtive to the SV model.

is further seen in the QQ-Plots[6]. Whilst both of these models provide significant increases in likelihood (insample) clearly neither model provides out-of-sample predictions that are superior to the SV model. This is seen both graphically in the QQ-plots and in likelihood values (out of sample) of -1657.4 and -1656.6 (for the $SV-t$ and $SV2$ models respectively).

The QQ-plots show that ASV volatility predictions appear to fit the lower tail of the distribution somewhat better than the standard SV model. Overall, the ASV model produces an out-of-sample likelihood of -1633.8, indicating that the ASV model produces more accurate forecasts than the SV model.

The first result of this Section is that the DNF procedure provides superior distributional forecasts when compared to the QML. Additionally, all three of the extended SV specifications provide significant insample gains over the SV model. It is only the ASV model however, that generates out of sample predictions that are superior to the SV model. This indicates that leverage is

[6] Whilst not reported here, a QQ-Plot of returns standardised by the $SV-t$ against returns standardised by the $SV2$ model is a perfect straight line.

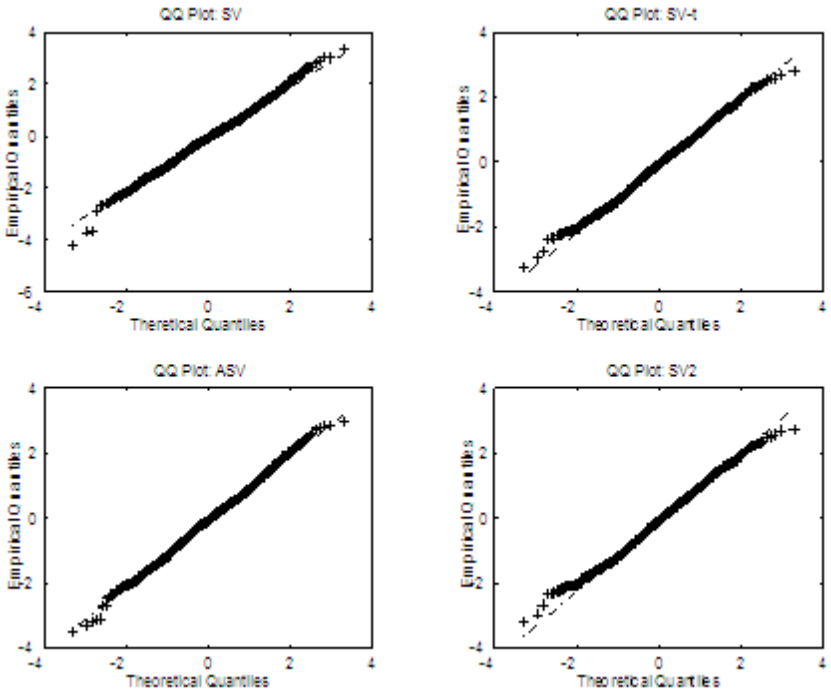

Figure 3: QQ-plots of returns standardised by SV, $SV-t$, ASV and $SV2$ volatility predictions.

a dominant feature of S&P500 returns.

7 CONCLUSION

The central contribution of this paper has been the proposal of a computationally efficient non-linear filtering algorithm. This algorithm is based on a fixed discretisation of the state space and permits maximum likelihood estimation of dynamic latent variable models such as the stochastic volatility model. Through simulation studies it has been shown that the DNF procedure generates accurate parameter estimates (in a computationally efficient manner) for the standard SV model.

It has also been shown that the DNF framework is flexible in that it can accommodate various extensions to the standard SV model. If the transition probabilities and likelihood function can be defined for a given latent variable process, the DNF estimation framework can be applied. In this paper asymmetric, heavy tailed and 2 factor SV models have been estimated using the DNF

framework. This is useful in that it enables a comparison of various features of the volatility process.

References

Alizadeh, S., Brandt, M., and Diebold, F. (2002), "Range based estimation of stochastic volatilty models" *The Journal of Finance*, 57, 1047-1091.

Black, F. (1976), "Studies in stock volatility changes", Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association, 177-181.

Campbell, J.Y. and Hentschel, L. (1992), "No news is good news: An asymmetric model of changing volatility in stock returns", *Journal of Financial Economics*, 31, 281-318.

Chernov, M., Gallant, R., Ghysels, E., and Tauchen, G. (2003), "Alternative models for stock price dynamics", *Journal of Econometrics*, 116, 225-257.

Chib, S, Nardari, F, Shephard, N. (2002), "Markov chain monte carlo methods for stochastic volatility models", *Journal of Econometrics*, 108, 281-316.

Danielsson, J. (1994a), "Stochastic volatility in asset prices: estimation with simulated maximum likelihood", *Journal of Econometrics*, 64, 375-400.

Danielsson, J. and Richard, J.F. (1993), "Accelerated Gaussian importance sampler with application to dynamic latent variable models", *Journal of Applied Econometrics*, 8, 153-173.

Fridman, M, and Harris L. (1998), "A maximum likelihood approach for non-gaussian stochastic volatility models", *Journal of Business and Economic Statistics,* 16, 284-291.

Gallant, R, and Tauchen, G. (1996), "Which moments to match?", *Econometric theory*, 12, 657-681.

Ghysels E, Harvey AC, and Renault E. (1996). Stochastic volatility, In Statistical Methods in Finance, Maddala GS, Rao CR (eds). North Holland: Amsterdam.

Harvey AC, Ruiz E, and Shephard N. (1994), "Multivariate stochastic variance models", *Review of Economic Studies,* 61: 247-264.

Harvey, A. and Shephard, N. (1996), "Estimation of an asymmeteric stochasic volatility model for asset returns" *Journal of Business and Eonomic Statistics*, 14, 429-434.

Jacquier E, Polson G, and Rossi PE. (1994), "Bayesian analysis of stochastic volatility models", *Journal of Business and Economic Statistics,* 12, 371-389.

Jacquier E, Polson G, and Rossi PE. (2004), "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, Journal of Econometrics", 122, 185-212.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic volatility: likelihood inference and comparison with ARCH models" *Review of Economic Studies*, 65, 361-393.

Kitagawa G. (1987), "Non-gaussian state-space modeling of non-stationary time series", *Journal of the American Statistical Association,* 82, 1032-1062.

Liesenfeld, R and Jung, R.C. (2000), "Stochastic volatility models: conditional normaility versus heav-ytailed distributions." *Journal of Applied Econometrics*, 15, 137-160.

Liesenfled, R. and Richard, JF. (2003), "Univariate and mulitvariate stochastic volatility models: estimation and diagnostics", *Journal of Empirical Finance*, 10, 505-531.

MacDonald IL, Zucchini W. (1997), Hidden Markov and Other Models for Discrete-valued Time Series. Chapman and Hall: London.

Martin DR, Raferty AE. (1987). "Robustness, computation and non-Euclidean models", *Journal of the American Statistical Association*, 82, 1044-1050.

Melino, A. and Turnbull, S. (1990), "Pricing foreign currency options with stochastic volatility", *Journal of Econometrics*, 45, 239-265.

Sandman, G., Koopman, S. J. (1998), "Estimation of stochastic volatility models via monte carlo maximum likelihood", *Journal of Econometrics*, 87, 271-301.

Taylor S (1982),"Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices, 1961-79, in Time Series Analysis : Theory and Practice 1, O.D. Anderson editor, North Holland, Amsterdam, 203-226.

Taylor S, 1986, Modelling Financial Time Series, Wiley, Chichester.

Watanabe T. (1999), "A non-linear filtering approach to stochastic volatility models with an application to daily stock returns", *Journal of Applied Econometrics,* 14, 101-121.

Watanabe, T., Asai, M. (2001), Stochastic volatility models with heavy-tailed distributions: a bayesian analysis, IMES Discussion paper series, 2001-E-17.

Yu, J. (2005), "On leverage in a stochastic volatility model", *Journal of Econometrics*, In Press.